Eukaryotic Transcription

Introduction

This lesson assumes that you have mastered prokaryotic transcription. If not, do that first. We're diving in. All of our references are to the 5' end of the coding strand, which is the same as the 5' end of the mRNA, except we have T's in DNA and U's in RNA. The template strand is almost never referenced. Explaining the flipping-a-flip-to-get-to-where-you-started is way more complicated than saying "*the coding DNA strand is the same as the transcribed mRNA strand, except T goes to U.*"

What Eukaryotic DNA Looks Like

Eukaryotic DNA has a **promoter region**. It consists of two consensus sequences. A **CAAT box at -70** and a **TATA box at -25**. There's a **+1 site**. That sounds an awful lot like prokaryotic DNA. Yep—so far this is the same as in prokaryotic, except for the **locations** of the consensus sequences (-70 and -25 instead of -35 and -10), and we've named the first promoter region the CAAT box (unnamed in prokaryotic).

The **transcription unit** starts at +1. It has a 5' UTR with a start sequence, ATG. There is **no code for a Shine-Dalgarno sequence** in a eukaryote. The transcription unit has a 3' UTR with a termination sequence. The 3' UTR doesn't have a GC-rich area to form a hairpin, and it doesn't have a TTTTTT sequence that will become a bunch of U's. Instead, the eukaryotic DNA termination sequence has the **poly(A) signal AATAAA** which will become AAUAAA on the mRNA.

The REALLY BIG DIFFERENCE in **eukaryotic DNA** is what's actually inside the **coding unit**. There's **one start translation** signal and **one stop translation** signal. Within those start and stop signals, there are **exons** and **introns**. Exons exist in the final mRNA and exit the nucleus. Introns don't. As a coding sequence of DNA, the entire thing gets transcribed, just as a polycystronic DNA coding sequence had its entire coding region transcribed. Where this gets messy is what happens to the mRNA, the actual focus of this lesson.

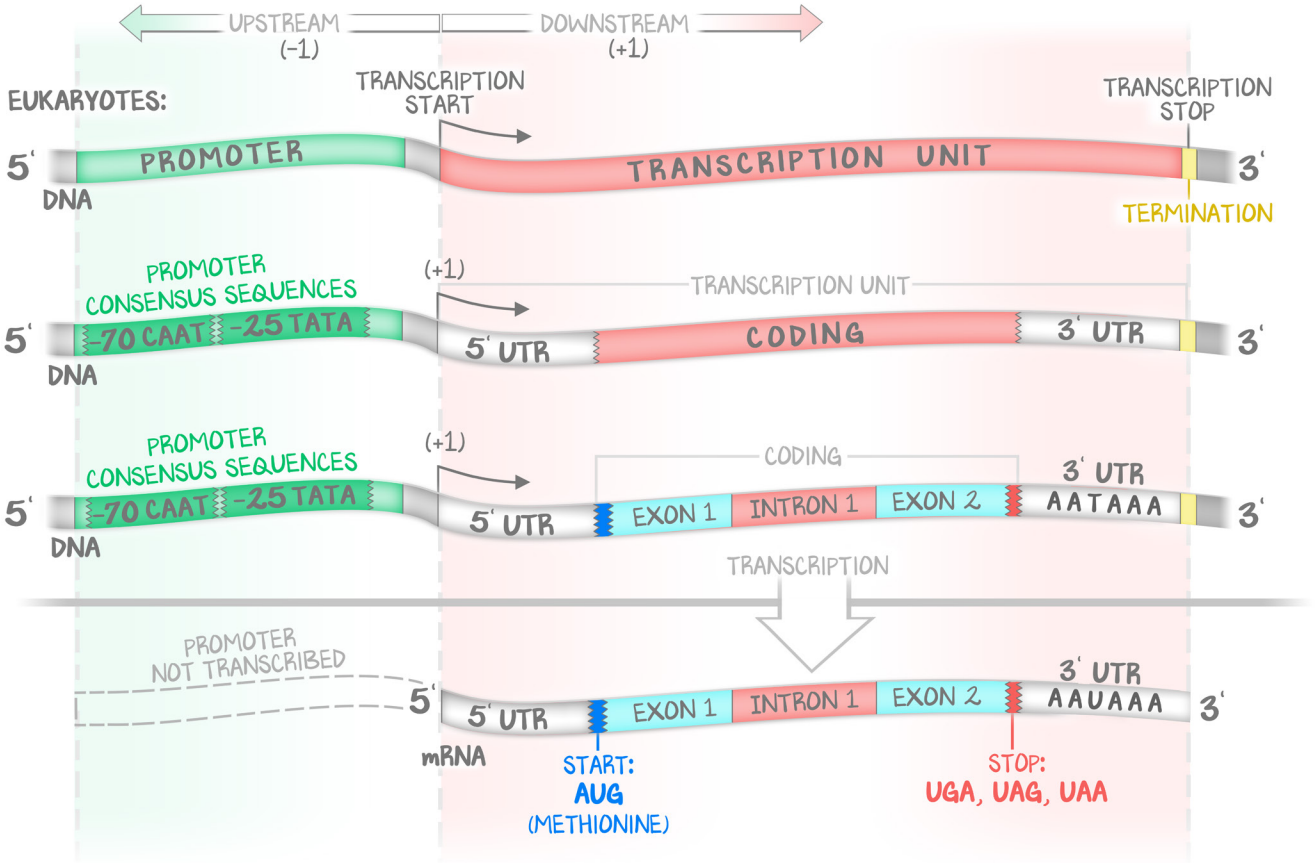

Figure 9.1: Eukaryotic DNA to mRNA
Using the same degree of depth in lesson #8, the details of eukaryotic DNA and hnRNA are shown. The major difference is the presence of introns, the absence of Shine-Dalgarno units, and a slight twist in the 3' tail.

What Eukaryotic hnRNA Looks Like

We've been lying. In the prokaryote, DNA is transcribed into mRNA. That was true. And in eukaryotes DNA will eventually become mRNA. But in reality, eukaryotic DNA will be transcribed and then modified into mRNA. The **primary transcript**, the thing that comes out of the RNA polymerase in prokaryotes, is prokaryotic mRNA. The primary transcript of eukaryotic DNA is hnRNA. **hnRNA** is the **in-processing** version of what will become mRNA. By the end of this section, you'll know what hnRNA looks like. By the end of the lesson, you'll know what eukaryotic mRNA looks like, too.

The **promoter region doesn't get transcribed**. It's absent from hnRNA (just as the promoter region is absent from prokaryotic mRNA).

The entire **transcription unit** from +1 to termination is coded into hnRNA. There's a 5' UTR region (**without a Shine-Dalgarno sequence**), a start signal of AUG, a stop signal of UAG/UAA/UGA, and 3' UTR region. The transcribed hnRNA has the coding for both exons and introns. The 3' UTR has the AAUAAA sequence, the **poly(A) addition sequence**.

The differences between eukaryotic hnRNA and prokaryotic mRNA so far are only the absence of a Shine-Dalgarno sequence and a poly(A) addition sequence instead of the GC-hairpin loop. If we called all the exons and introns "coding region" those two things (Shine-Dalgarno and poly(A) addition) would be the only differences*.

Preprocessed hnRNA does look a lot like prokaryotic mRNA. But the processing makes eukaryotic mRNA look more different than prokaryotic mRNA. Here's how.

Processing of Eukaryotic hnRNA

There are three processing events: methyl-guanine cap, poly(A) tail, and splicing.

The addition of the **methyl-guanine cap** to the **5' end** is the first processing event, and the only processing event that occurs at the same time as transcription. It's called **co-transcription**, and it stabilizes **the final mRNA** to allow it to exit the nucleus.

The other two processing steps are post-transcriptional.

The second processing event is the **addition of a poly(A) tail** to the **3' end**. This event has two steps. First, **endonucleolytic cleavage** removes the AAUAAA piece at the 3' end. Second, **poly(A) polymerase** recognizes the site of endonucleolytic cleavage, and places a poly-Adenine tail, an extremely long sequence of **hundreds of adenines** in a row. Exonucleases of the cell and nucleus eat the RNA at the 3' end. The poly(A) tail will be degraded first. Then, once it's gone, the actual mRNA will be degraded. The longer the poly(A) tail, the longer the transcript will be present and the more protein will be made. This constitutes not only a necessary addition for export to the cytoplasm, but also a mechanism for regulating gene expression (i.e., how much protein will be made is dependent on the length of the poly(A) tail).

The third processing event is called **splicing**. Splicing is the removal of the introns.

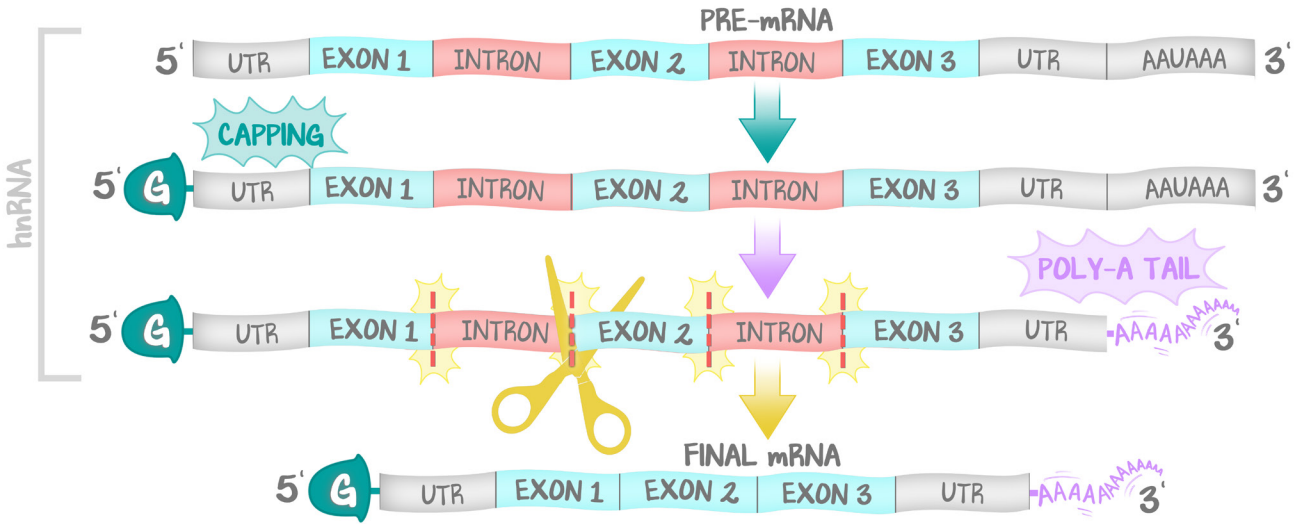

Figure 9.2: hnRNA Processing to Mature mRNA
hnRNA is the premature eukaryotic RNA that must undergo the addition of a methyl-guanine cap, a poly(A) tail, and removal of introns prior to becoming mature mRNA that will be allowed to pass through the nuclear pore.

Splicing

The hnRNA is going 5' to 3', left to right. The first exon starts with its 5' left and 3' right. The second exon starts with its 5' left and 3' right. In between is an intron. Screw that intron. THROW HIM OUT. A protein called a spliceosome goes "*clip, clip*" and the intron is gone. Bye, intron. So now there's a 5'→3' Exon, a gaping hole, and another 5'→3' exon. That same spliceosome grabs the 3' end of exon on the left, the 5' end of the exon on the right, and *snap*—hooks them together.

The **spliceosome** (snRNP, aka snurp) is a combination of snRNA and protein. The spliceosome is responsible for **removal of the intron**. The spliceosome is responsible for connecting the exons together.

See how we didn't use "donor" or "acceptor," and how we maintained that visual of 5' left, 3' right? That's about as much as you need to know. We know that biochemistry PhDs will be disgusted with that explanation. For everyone else: just be sure **always to orient your hnRNA 5'-start-left and 3'-end-right**. In other words, how we've done it in every lesson. Biochemistry class might have tried to screw you up by inverting the sequences, and the test CAN do that, but since Step 1 isn't about curveballs, it's virtually guaranteed everything will be 5' to 3'.

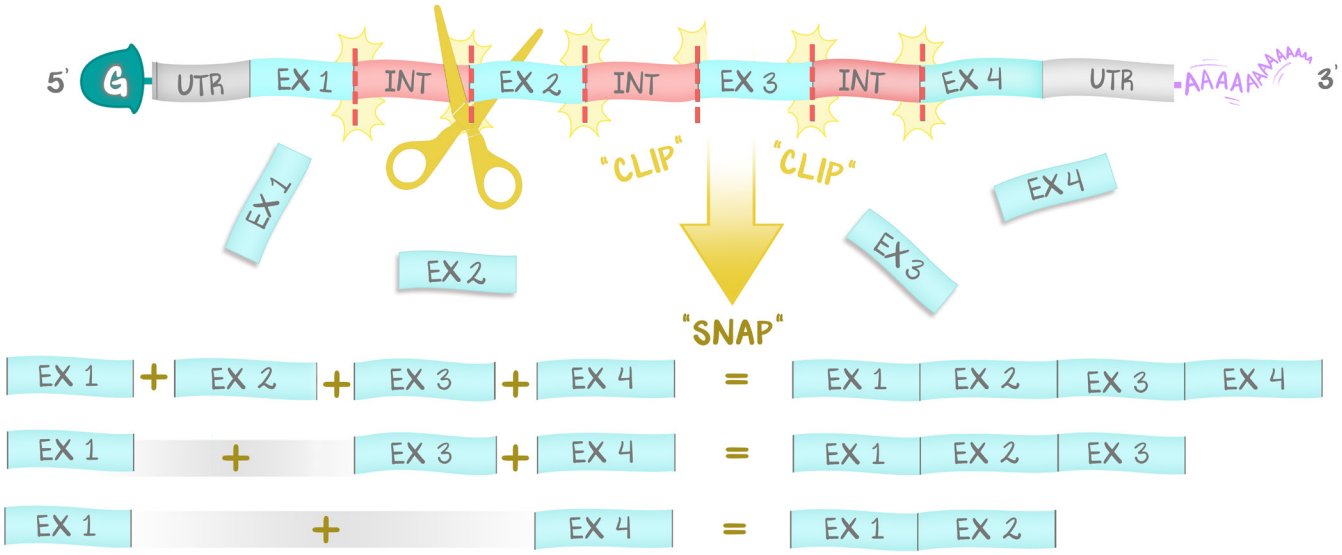

Figure 9.3: Alternative Splicing
Depending on which exons are removed, one transcription unit can produce variable mRNA sequences.

Export to Cytoplasm

The nucleus is smart. It knows eukaryotic RNA requires processing. A **nuclear pore complex** will deny exit of any mRNA that doesn't pass its **three** necessary checks. The nuclear pore will allow export only if there is the 5' methyl-guanine cap, the poly(A) tail bound to poly(A)-binding protein (the binding protein tells the pore it's good, and binding protein can only bind to the completed poly(A) tail), and the splice-binding proteins at the splice sites.

The finished product of processing and splicing is called **mRNA**. It's sent to cytoplasm.

EXTRA: Alternative Splicing

OH GNOES! IT GETS HARDER!?

No. It doesn't. So there's a coding region with four (4) exons now. Exon 1 has its 5' to the left, 3' to right. Exon 2 has its 5' to the left, 3' to the right. Exon 3 has its . . . you get it. Each exon has introns in between.

Spliceosome'd!

All the introns are gone. Exon1____Exon2____Exon3_____Exon4. This is why scientists make up complicated language. Because the spliceosome doesn't just "close the gap," as I implied it did above. The spliceosome can take the beginning of any one exon and "snap" it to the end of any other exon. If it does that, any exons in between are eliminated. So it's the same thing, except that spliceosomes can jump over an exon, eliminating it.

The **reason** for alternative splicing is more interesting than **how it happens**. The primary transcript, the rigid strict code of the pre-mRNA, is just a copy of the DNA. Since there is more than one type of cell in the body, wouldn't it be nifty if the same code could do different things in different cells? That's what happens. In one cell, the primary transcript of all-the-exons-and-all-the-introns might be spliced to Ex1-Ex2-Ex4 and the protein made by that will do something. In another cell, the primary transcript of all-the-exons-and-all-the-introns might be spliced to Ex1-Ex3-Ex4 and the protein it makes is different. And cooler still, even within the same cell, the same primary transcript might be processed differently at different life stages of the cell.

For example, the mature naïve B lymphocyte produces immunoglobulin that sticks to the cell surface. This is called surface-bound membrane-protein. It's on the surface because it's detecting antigen. If it detects antigen, it gets internalized, processed, and shown to nearby T cells. Endocytosis is that immunoglobulin's purpose. When activated, that B lymphocyte wants to make that same immunoglobulin but instead of sticking on its surface, it wants to secrete the immunoglobulin. That secreted immunoglobulin still detects antigen, but its free part, the part that used to be stuck on the B cell surface, now can interact with other cells. The switch from membrane-bound immunoglobulin to secreted immunoglobulin is managed by alternative splicing.